Procédé d'analyse d'informations de fréquence fondamentale et procédé et système de conversion de voix mettant en œuvre un tel procédé d'analyse

La présente invention concerne un procédé d'analyse d'informations de fréquence fondamentale contenues dans des échantillons vocaux, et un procédé et un système de conversion de voix mettant en œuvre ce procédé d'analyse.

Suivant la nature des sons à émettre, la production de la parole et notamment des sons voisés, peut faire intervenir la vibration des cordes vocales, ce qui se manifeste par la présence dans le signal de parole, d'une structure périodique de période fondamentale dont l'inverse est appelé fréquence fondamentale ou "pitch".

Dans certaines applications, tels que la conversion de voix, le rendu auditif est primordial et pour obtenir une qualité acceptable, il convient de bien maîtriser les paramètres liés à la prosodie et parmi ces derniers, la fréquence fondamentale.

Ainsi, il existe aujourd'hui de nombreux procédés d'analyse des informations de fréquence fondamentale contenues dans des échantillons vocaux.

Ces analyses permettent de déterminer et de modéliser des caractéristiques de la fréquence fondamentale. Par exemple, il existe des procédés permettant de déterminer la pente, ou encore une échelle d'amplitude de la fréquence fondamentale sur l'ensemble d'une base de données d'échantillons vocaux.

La connaissance de ces paramètres permet d'effectuer des modifications de signaux de parole, par exemple par des mises à l'échelle de fréquence fondamentale entre des locuteurs source et cible, de manière à respecter globalement la moyenne et la variation de la fréquence fondamentale du locuteur cible.

Cependant, ces analyses ne permettent d'obtenir que des représentations globales et pas de représentations paramétrables de la fréquence fondamentale et ne sont donc pas pertinentes notamment pour des locuteurs dont les styles d'élocution sont différents.

Le but de la présente invention est de remédier à ce problème, en définissant un procédé d'analyse d'informations de fréquence fondamentale d'échantillons vocaux, permettant la définition d'une représentation paramétrable de la fréquence fondamentale.

A cet effet, la présente invention a pour objet un procédé d'analyse d'informations de fréquence fondamentale contenues dans des échantillons vocaux, caractérisé en ce qu'il comporte au moins :

- une étape d'analyse des échantillons vocaux regroupés en trames pour obtenir, pour chaque trame d'échantillons, des informations relatives au spectre et des informations relatives à la fréquence fondamentale;

- une étape de détermination d'un modèle représentant les caractéristiques communes de spectre et de fréquence fondamentale de tous les échantillons; et

- une étape de détermination, à partir de ce modèle et des échantillons vocaux, d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre.

Suivant d'autres caractéristiques de ce procédé d'analyse :

- ladite étape d'analyse est adaptée pour délivrer lesdites informations relatives au spectre sous la forme de coefficients cepstraux ;

- ladite étape d'analyse comporte :

- une sous-étape de modélisation des échantillons vocaux selon une somme d'un signal harmonique et d'un signal de bruit ;

- une sous-étape d'estimation de paramètres de fréquence et au moins de la fréquence fondamentale des échantillons vocaux ;

- une sous-étape d'analyse synchronisée de chaque trame d'échantillons sur sa fréquence fondamentale ; et

- une sous-étape d'estimation des paramètres de spectre de chaque trame d'échantillons ;

- il comporte en outre une étape de normalisation de la fréquence fondamentale de chaque trame d'échantillons par rapport à la moyenne des fréquences fondamentales des échantillons analysés ;

- ladite étape de détermination d'un modèle correspond à la détermination d'un modèle par mélange de densités gaussiennes ;

- ladite étape de détermination d'un modèle comprend :

- une sous-étape de détermination d'un modèle correspondant à un mélange de densités gaussiennes; et

- une sous-étape d'estimation des paramètres du mélange de densités gaussiennes à partir de l'estimation du maximum de vraisemblance en-

tre les informations de spectre et de fréquence fondamentale des échantillons et du modèle ;

- ladite étape de détermination d'une fonction de prédiction est réalisée à partir d'un estimateur de la réalisation de la fréquence fondamentale sachant les informations de spectre des échantillons ;

- ladite étape de détermination de la fonction de prédiction de la fréquence fondamentale comprend une sous-étape de détermination de l'espérance conditionnelle de la réalisation de la fréquence fondamentale sachant les informations de spectre à partir de la probabilité a posteriori que les informations de spectre soient obtenues à partir du modèle, l'espérance conditionnelle formant ledit estimateur.

L'invention a également pour objet un procédé de conversion d'un signal vocal prononcé par un locuteur source en un signal vocal converti dont les caractéristiques ressemblent à celles d'un locuteur cible, comportant au moins :

- une étape de détermination d'une fonction de transformation de caractéristiques spectrales du locuteur source en caractéristiques spectrales du locuteur cible, réalisée à partir d'échantillons vocaux du locuteur source et du locuteur cible; et

- une étape de transformation des informations de spectre du signal de voix du locuteur source à convertir à l'aide de ladite fonction de transformation,

caractérisé en ce qu'il comporte en outre :

- une étape de détermination d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre pour le locuteur cible, ladite fonction de prédiction étant obtenue à l'aide d'un procédé d'analyse tel que défini précédemment ; et

- une étape de prédiction de la fréquence fondamentale du signal de voix à convertir par l'application de ladite fonction de prédiction de la fréquence fondamentale auxdites informations de spectres transformés du signal de voix du locuteur source.

Suivant d'autres caractéristiques de ce procédé de conversion :

- ladite étape de détermination d'une fonction de transformation est réalisée à partir d'un estimateur de la réalisation des caractéristiques spectrales cibles sachant les caractéristiques spectrales source ;

- ladite étape de détermination d'une fonction de transformation comporte :

- une sous-étape de modélisation des échantillons vocaux source et cible selon un modèle de somme d'un signal harmonique et d'un signal de bruit ;

- une sous-étape d'alignement entre les échantillons source et cible; et

- une sous-étape de détermination de ladite fonction de transformation à partir du calcul de l'espérance conditionnelle de la réalisation des caractéristiques spectrales cibles sachant la réalisation des caractérisations spectrales sources, l'espérance conditionnelle formant ledit estimateur.

- ladite fonction de transformation est une fonction de transformation de l'enveloppe spectrale ;

- il comporte en outre une étape d'analyse du signal de voix à convertir adaptée pour délivrer lesdites informations relatives au spectre et à la fréquence fondamentale ;

- il comporte en outre une étape de synthèse permettant de former un signal de voix converti à partir au moins des informations de spectre transformées et des informations de fréquence fondamentale prédites.

L'invention a encore pour objet un système de conversion d'un signal vocal prononcé par un locuteur source en un signal vocal converti dont les caractéristiques ressemblent à celles d'un locuteur cible, système comportant au moins :

- des moyens de détermination d'une fonction de transformation de caractéristiques spectrales du locuteur source en caractéristiques spectrales du locuteur cible, recevant en entrée des échantillons vocaux du locuteur source et du locuteur cible ; et

- des moyens de transformation des informations de spectre du signal de voix du locuteur source à convertir par l'application de ladite fonction de transformation délivrée par les moyens,

caractérisé en ce qu'il comporte en outre :

- des moyens de détermination d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre

pour le locuteur cible, adaptés pour la mise en œuvre d'un procédé d'analyse, à partir d'échantillons vocaux du locuteur cible ; et

- des moyens de prédiction de la fréquence fondamentale dudit signal de voix à convertir, par l'application de ladite fonction de prédiction déterminée par lesdits moyens de détermination d'une fonction de prédiction auxdites informations de spectre transformé délivrées par lesdits moyens de transformation.

Suivant d'autres caractéristiques de ce système :

- il comporte en outre :

- des moyens d'analyse du signal de voix à convertir, adaptés pour délivrer en sortie des informations relatives au spectre et à la fréquence fondamentale du signal de voix à convertir ; et

- des moyens de synthèse permettant de former un signal de voix converti à partir au moins des informations de spectre transformé délivrées par les moyens et des informations de fréquence fondamentale prédites délivrées par les moyens;

- lesdits moyens de détermination d'une fonction de transformation sont adaptés pour délivrer une fonction de transformation de l'enveloppe spectrale ;

- il est adapté pour la mise en œuvre d'un procédé de conversion de voix tel que défini précédemment.

L'invention sera mieux comprise à la lecture de la description qui va suivre, donnée uniquement à titre d'exemple et faite en se référant aux dessins annexés, sur lesquels :

- la Fig.1 est un organigramme d'un procédé d'analyse selon l'invention ;

- la Fig.2 est un organigramme d'un procédé de conversion de voix mettant en œuvre le procédé d'analyse de l'invention ; et

- la Fig.3 est un schéma bloc fonctionnel d'un système de conversion de voix, permettant la mise en œuvre du procédé de l'invention décrit à la figure 2.

Le procédé de l'invention représenté sur la figure 1, est mis en œuvre à partir d'une base de données d'échantillons vocaux contenant des séquences de parole naturelle.

Le procédé débute par une étape 2 d'analyse des échantillons en les regroupant par trame, afin d'obtenir pour chaque trame d'échantillons, des informations relatives au spectre et notamment à l'enveloppe spectrale et des informations relatives à la fréquence fondamentale.

Dans le mode de réalisation décrit, cette étape 2 d'analyse est basée sur l'utilisation d'un modèle d'un signal sonore sous la forme d'une somme d'un signal harmonique avec un signal de bruit selon un modèle communément appelé "HNM" (en anglais : Harmonic plus Noise Model).

En outre, le mode de réalisation décrit est fondé sur une représentation de l'enveloppe spectrale par le cepstre discret.

En effet, une représentation cepstrale permet de séparer, dans le signal de parole, la composante relative au conduit vocal de la composante résultant de la source, correspondant aux vibrations des cordes vocales et caractérisée par la fréquence fondamentale.

Ainsi, l'étape 2 d'analyse comporte une sous-étape 4 de modélisation de chaque trame de signal vocal en une partie harmonique représentant la composante périodique du signal, constituée d'une somme de L sinusoïdes harmoniques d'amplitude A_l et de phase ϕ_l, et d'une partie bruitée représentant le bruit de friction et la variation de l'excitation glottale.

On peut ainsi écrire :

$$s(n)=h(n)+b(n)$$

avec $$h(n)= \sum_{l=1}^{L} A_l(n)\cos(\phi_l(n))$$

Le terme h(n) représente donc l'approximation harmonique du signal s(n).

L'étape 2 comporte ensuite une sous-étape 5 d'estimation pour chaque trame, de paramètres de fréquence et notamment de la fréquence fondamentale, par exemple au moyen d'une méthode d'autocorrélation.

De manière classique, cette analyse HNM délivre la fréquence maximale de voisement. En variante, cette fréquence peut être fixée arbitrairement ou être estimée par d'autres moyens connus.

Cette sous-étape 5 est suivie d'une sous-étape 6 d'analyse synchronisée de chaque trame sur sa fréquence fondamentale, qui permet d'estimer les paramètres de la partie harmonique ainsi que les paramètres du bruit du signal.

Dans le mode de réalisation décrit, cette analyse synchronisée correspond à la détermination des paramètres des harmoniques par minimisation d'un critère de moindres carrés pondérés entre le signal complet et sa décomposition harmonique correspondant dans le mode de réalisation décrit, au signal de bruit estimé. Le critère noté E est égal à :

$$E = \sum_{n=-T_i}^{T_i} w^2(n)(s(n)-h(n))^2$$

Dans cette équation, w (n) est la fenêtre d'analyse et T_i est la période fondamentale de la trame courante.

Ainsi, la fenêtre d'analyse est centrée autour de la marque de la période fondamentale et a pour durée deux fois cette période.

L'étape 2 d'analyse comporte enfin une sous-étape 7 d'estimation des paramètres des composantes de l'enveloppe spectrale du signal en utilisant par exemple une méthode de cepstre discret régularisé et une transformation en échelle de Bark pour reproduire le plus fidèlement possible les propriétés de l'oreille humaine.

Ainsi, l'étape 2 d'analyse délivre, pour chaque trame de rang n d'échantillons de signal de parole, un scalaire noté x_n comprenant des informations de fréquence fondamentale et un vecteur noté y_n comprenant des informations de spectre sous la forme d'une séquence de coefficients cepstraux.

Avantageusement, l'étape 2 d'analyse est suivie par une étape 10 de normalisation de la valeur de la fréquence fondamentale de chaque trame par rapport à la fréquence fondamentale moyenne afin de remplacer pour chaque trame d'échantillons vocaux, la valeur de la fréquence fondamentale par une valeur de fréquence fondamentale normalisée selon la formule suivante :

$$F_{log} = \log\left(\frac{Fo}{F_o^{moy}}\right)$$

Dans cette formule, F_o^{moy} correspond à la moyenne des valeurs des fréquences fondamentales sur toute la base de données analysée.

Cette normalisation permet de modifier l'échelle des variations des scalaires de fréquence fondamentale afin de la rendre cohérente avec l'échelle des variations des coefficients cepstraux.

L'étape 10 de normalisation est suivie d'une étape 20 de détermination d'un modèle représentant les caractéristiques communes de cepstre et de fréquence fondamentale de tous les échantillons analysés.

Dans le mode de réalisation décrit, il s'agit d'un modèle probabiliste de la fréquence fondamentale et du cepstre discret, selon un modèle de mélange de densités gaussiennes couramment noté "GMM", dont les paramètres sont estimés à partir de la densité jointe de la fréquence fondamentale normalisée et du cepstre discret.

De manière classique, la densité de probabilité d'une variable aléatoire notée de manière générale p(z), suivant un modèle de mélange de densités gaussiennes GMM s'écrit mathématiquement de la manière suivante :

$$p(z)=\sum_{i=1}^{Q}\alpha_i=N(z,\mu_i,\Sigma_i)$$

avec $$\sum_{i=1}^{Q}\alpha_{i,}=1,\ o\leq\alpha_i\leq 1$$

Dans cette formule, $N(z\ ;\ \mu_i\ ;\ \Sigma_i)$ est la densité de probabilité de la loi normale de moyenne μ_i et de matrice de covariance Σ_i et les coefficients α_i sont les coefficients du mélange.

Ainsi, le coefficient α_i correspond à la probabilité a priori que la variable aléatoire z soit générée par la $i^{ème}$ gaussienne du mélange.

De manière plus particulière, l'étape 20 de détermination du modèle comporte une sous-étape 22 de modélisation de la densité jointe entre le cepstre noté y et la fréquence fondamentale normalisée notée x, de sorte que :

$$p(z) = p(y,x),\ \text{où}\ z = \begin{pmatrix} y \\ x \end{pmatrix}$$

Dans ces équations, $x = [x_1, x_2, \ldots x_N]$ correspond à la séquence des scalaires contenant les informations de fréquence fondamentale normalisée pour N trames d'échantillons vocaux et $y = [y_1, y_2, \ldots y_N]$, correspond à la séquence des vecteurs de coefficients cepstraux correspondants.

L'étape 20 comporte ensuite une sous-étape 24 d'estimation de paramètres GMM (α, μ, Σ) de la densité p(z). Cette estimation peut être réalisée, par exemple, à l'aide d'un algorithme classique de type dit "EM" (Expectation – Maximisation), correspondant à une méthode itérative conduisant à l'obtention

d'un estimateur de maximum de vraisemblance entre les données des échantillons de parole et le modèle de mélange de gaussienne.

La détermination des paramètres initiaux du modèle GMM est obtenue à l'aide d'une technique classique de quantification vectorielle.

L'étape 20 de détermination de modèle délivre ainsi les paramètres d'un mélange de densités gaussiennes représentatifs des caractéristiques communes des spectres, représentées par les coefficients cepstraux, et des fréquences fondamentales des échantillons vocaux analysés.

Le procédé comporte ensuite une étape 30 de détermination, à partir du modèle et des échantillons vocaux, d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations de spectre fournies par le cepstre du signal.

Cette fonction de prédiction est déterminée à partir d'un estimateur de la réalisation de la fréquence fondamentale étant donné le cepstre des échantillons vocaux, formé dans le mode de réalisation décrit, par l'espérance conditionnelle.

Pour cela, l'étape 30 comporte une sous-étape 32 de détermination de l'espérance conditionnelle de la fréquence fondamentale sachant les informations relatives au spectre fournies par le cepstre. L'espérance conditionnelle est notée F(y) et est déterminée à partir des formules suivantes :

$$F(y)=E[x\,|\,y]=\sum_{i=1}^{Q}P_i(y)[\mu_i^x+\Sigma_i^{xy}(\Sigma_i^{yy})^{-1}(y-\mu_i^y)]$$

avec $$P_i(y)=\frac{\alpha_i N(y,\mu_i^y,\Sigma_i^{yy})}{\sum_{j=1}^{Q}\alpha_j N(y,\mu_j^y,\Sigma_j^{yy})}$$

avec $$\Sigma_i=\begin{bmatrix}\Sigma_i^{yy} & \Sigma_i^{yx}\\ \Sigma_i^{xy} & \Sigma_i^{xx}\end{bmatrix} \text{ et } \mu_i=\begin{bmatrix}\mu_i^x\\ \mu_i^y\end{bmatrix}$$

Dans ces équations, $P_i(y)$ correspond à la probabilité a posteriori que le vecteur y de cepstre soit généré par la $i^{ème}$ composante du mélange de gaussiennes du modèle, défini lors de l'étape 20 par la matrice de covariance Σ_i et la loi normale μ_i.

La détermination de l'espérance conditionnelle permet ainsi d'obtenir la fonction de prédiction de la fréquence fondamentale à partir des informations de cepstre.

En variante, l'estimateur mis en œuvre lors de l'étape 30 peut être un critère de maximum a posteriori, dit "MAP" et correspondant à la réalisation du calcul de l'espérance uniquement pour le modèle représentant le mieux le vecteur source.

Il apparaît donc que le procédé d'analyse de l'invention permet, à partir du modèle et des échantillons vocaux, d'obtenir une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations de spectre fournies, dans le mode de réalisation décrit, par le cepstre.

Une telle fonction de prédiction permet ensuite de déterminer la valeur de la fréquence fondamentale pour un signal de parole, uniquement à partir d'informations de spectre de ce signal, permettant ainsi une prédiction pertinente de la fréquence fondamentale notamment pour des sons qui ne sont pas dans les échantillons vocaux analysés.

En référence à la figure 2, on va maintenant décrire l'utilisation d'un procédé d'analyse selon l'invention dans le cadre de la conversion de voix.

La conversion de voix consiste à modifier le signal vocal d'un locuteur de référence appelé " locuteur source " de façon que le signal produit semble avoir été prononcé par un autre locuteur nommé " locuteur cible".

Ce procédé est mis en œuvre à partir d'une base de données d'échantillons vocaux prononcés par le locuteur source et le locuteur cible.

De manière classique, un tel procédé comporte une étape 50 de détermination d'une fonction de transformation des caractéristiques spectrales des échantillons vocaux du locuteur source pour les faire ressembler aux caractéristiques spectrales des échantillons vocaux du locuteur cible.

Dans le mode de réalisation décrit, cette étape 50 est basée sur une analyse de type HNM permettant de déterminer les relations existantes entre les caractéristiques de l'enveloppe spectrale des signaux de parole des locuteurs source et cible.

Pour cela, il est nécessaire de disposer d'enregistrements vocaux source et cible correspondant à la réalisation acoustique de la même séquence phonétique.

L'étape 50 comporte une sous-étape 52 de modélisation des échantillons vocaux selon un modèle HNM, de somme de signaux harmoniques et de bruit.

La sous-étape 52 est suivie d'une sous-étape 54 d'alignement entre les signaux source et cible à l'aide par exemple d'un algorithme classique d'alignement dit "DTW" (en anglais " Dynamic Time Warping").

L'étape 50 comporte ensuite une sous-étape 56 de détermination d'un modèle tel qu'un modèle de type GMM représentant les caractéristiques communes des spectres des échantillons vocaux des locuteurs source et cible.

Dans le mode de réalisation décrit, on utilise un modèle GMM à 64 composantes et un unique vecteur contenant les paramètres cepstraux de la source et de la cible, de sorte que l'on peut définir une fonction de transformation spectrale correspondant à un estimateur de la réalisation des paramètres spectraux de cible notés t, sachant les paramètres spectraux de source notés s.

Dans le mode de réalisation décrit, cette fonction de transformation notée F(s) se note sous la forme d'une espérance conditionnelle obtenue par la formule suivante :

$$F(s)=E[t \mid s]=\sum_{i=1}^{Q} P_i(s)[\mu_i^t+\Sigma_i^{ts}(\Sigma_i^{ss})^{-1}(s-\mu_i^s)]$$

avec $$P_i(s)=\frac{\alpha_i N(s,\mu_i^s,\Sigma_i^{ss})}{\sum_{j=1}^{Q}\alpha_i N(t,\mu_j^s,\Sigma_j^{ss})}$$

avec $$\Sigma_i=\begin{bmatrix}\Sigma_i^{ss} & \Sigma_i^{st}\\ \Sigma_i^{ts} & \Sigma_i^{tt}\end{bmatrix} \text{ et } \mu_i=\begin{bmatrix}\mu_i^s\\ \mu_i^t\end{bmatrix}$$

La détermination précise de cette fonction est obtenue par la maximisation de la vraisemblance entre les paramètres de la source et de la cible, obtenue par un algorithme de type EM.

En variante, l'estimateur peut être formé d'un critère de maximum a posteriori.

La fonction ainsi définie permet donc de modifier l'enveloppe spectrale d'un signal de parole issue du locuteur source afin de la faire ressembler à l'enveloppe spectrale du locuteur cible.

Préalablement à cette maximisation, les paramètres du modèle GMM représentant les caractéristiques spectrales communes de la source et de la cible sont initialisés, par exemple, à l'aide d'un algorithme de quantification vectorielle.

Parallèlement, le procédé d'analyse de l'invention est mis en œuvre lors d'une étape 60 d'analyse des seuls échantillons vocaux du locuteur cible.

Ainsi que cela a été décrit à la référence à la figure 1, l'étape 60 d'analyse selon l'invention permet d'obtenir, pour le locuteur cible, une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations de spectres.

Le procédé de conversion comporte ensuite une étape 65 d'analyse d'un signal de voix à convertir prononcé par le locuteur source, lequel signal à convertir est différent des signaux vocaux utilisés lors des étapes 50 et 60.

Cette étape d'analyse 65 est réalisée, par exemple, à l'aide d'une décomposition selon le modèle HNM permettant de délivrer des informations de spectre sous la forme de coefficients cepstraux, des informations de fréquence fondamentale ainsi que des informations de phase et de fréquence maximale de voisement.

Cette étape 65 est suivie d'une étape 70 de transformation des caractéristiques spectrales du signal de voix à convertir par l'application de la fonction de transformation déterminée à l'étape 50, aux coefficients cepstraux définis lors de l'étape 65.

Cette étape 70 permet notamment la modification de l'enveloppe spectrale du signal de voix à convertir.

A l'issue de l'étape 70, chaque trame d'échantillons du signal à convertir du locuteur source est ainsi associée à des informations spectrales transformées dont les caractéristiques sont similaires aux caractéristiques spectrales des échantillons du locuteur cible.

Le procédé de conversion comporte ensuite une étape 80 de prédiction de la fréquence fondamentale pour les échantillons vocaux du locuteur source, par l'application de la fonction de prédiction déterminée selon le procédé de l'invention lors de l'étape 60, aux seules informations spectrales transformées associées au signal de voix à convertir du locuteur source.

En effet, les échantillons vocaux du locuteur source étant associés à des informations spectrales transformées dont les caractéristiques sont similaires

à celles du locuteur cible, la fonction de prédiction définie lors de l'étape 60 permet d'obtenir une prédiction pertinente de la fréquence fondamentale.

De manière classique, le procédé de conversion comporte ensuite une étape 90 de synthèse du signal de sortie réalisée, dans l'exemple décrit, par une synthèse de type HNM qui délivre directement le signal de voix converti à partir des informations d'enveloppe spectrale transformées délivrées par l'étape 70, des informations de fréquence fondamentale prédites issues de l'étape 80 et des informations de phase et de fréquence maximale de voisement délivrées par l'étape 65.

Le procédé de conversion mettant en œuvre le procédé d'analyse de l'invention, permet ainsi d'obtenir une conversion de voix réalisant des modifications de spectres ainsi qu'une prédiction de fréquence fondamentale, de manière à obtenir un rendu auditif de bonne qualité.

Notamment, l'efficacité d'un tel procédé peut être évaluée à partir d'échantillons vocaux identiques prononcés par le locuteur source et le locuteur cible.

Le signal vocal prononcé par le locuteur source est converti à l'aide du procédé tel que décrit et la ressemblance du signal converti avec le signal prononcé par le locuteur cible, est évaluée.

Par exemple, cette ressemblance est calculée sous la forme d'un ratio entre la distance acoustique séparant le signal converti du signal cible et la distance acoustique séparant le signal cible du signal source.

En calculant la distance acoustique à partir des coefficients cepstraux ou du spectre d'amplitude des signaux obtenu à l'aide de ces coefficients cepstraux, le ratio obtenu pour un signal converti à l'aide du procédé de l'invention est de l'ordre de 0,3 à 0,5.

Sur la figure 3, on a représenté un schéma bloc fonctionnel d'un système de conversion des voix mettant en œuvre le procédé décrit en référence à la figure 2.

Ce système utilise en entrée une base de données 100 d'échantillons vocaux prononcés par le locuteur source et une base de données 102 contenant au moins les mêmes échantillons vocaux prononcés par le locuteur cible.

Ces deux bases de données sont utilisées par un module 104 de détermination d'une fonction de transformation de caractéristiques spectrales du locuteur source en caractéristiques spectrales du locuteur cible.

Ce module 104 est adapté pour la mise en œuvre de l'étape 50 du procédé tel que décrit en référence à la figure 2 et permet donc la détermination d'une fonction de transformation de l'enveloppe spectrale.

Par ailleurs, le système comporte un module 106 de détermination d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre. Le module 106 reçoit pour cela en entrée les échantillons vocaux du seul locuteur cible, contenus dans la base de données 102.

Le module 106 est adapté pour la mise en œuvre de l'étape 60 du procédé décrit en référence à la figure 2 et correspondant au procédé d'analyse de l'invention tel que décrit en référence à la figure 1.

Avantageusement, la fonction de transformation délivrée par le module 104 et la fonction de prédiction délivrée par le module 106, sont mémorisées en vue d'une utilisation ultérieure.

Le système de conversion de voix reçoit en entrée un signal de voix 110 correspondant à un signal de parole prononcé par le locuteur source et destiné à être converti.

Le signal 110 est introduit dans un module 112 d'analyse du signal, mettant en œuvre, par exemple, une décomposition de type HNM et permettant de dissocier des informations de spectre du signal 110 sous la forme de coefficients cepstraux et d'informations de fréquence fondamentale. Le module 112 délivre également des informations de phase et de fréquence maximale de voisement obtenues par l'application du modèle HNM.

Le module 112 met donc en œuvre l'étape 65 du procédé décrit précédemment.

Eventuellement cette analyse peut être faite au préalable et les informations sont stockées pour être utilisées ultérieurement.

Les coefficients cepstraux délivrés par le module 112, sont ensuite introduits dans un module 114 de transformation adapté pour appliquer la fonction de transformation déterminée par le module 104.

Ainsi, le module 114 de transformation met en œuvre l'étape 70 du procédé décrit en référence à la figure 2 et délivre des coefficients cepstraux transformés dont les caractéristiques sont similaires aux caractéristiques spectrales du locuteur cible.

Le module 114 réalise ainsi une modification de l'enveloppe spectrale du signal de voix 110.

Les coefficients cepstraux transformés délivrés par le module 114, sont ensuite introduits dans un module 116 de prédiction de la fréquence fondamentale adaptés pour mettre en œuvre la fonction de prédiction déterminée par le module 106.

Ainsi, le module 116 met en œuvre l'étape 80 du procédé décrit en référence à la figure 2 et délivre en sortie des informations de fréquence fondamentale prédites à partir uniquement des informations de spectre transformées.

Le système comporte ensuite un module 118 de synthèse recevant en entrée les coefficients cepstraux transformés issus du module 114 et correspondant à l'enveloppe spectrale, les informations de fréquence fondamentale prédites issues du module 116, et les informations de phase et de fréquence maximale de voisement délivrées par le module 112.

Le module 118 met ainsi en œuvre l'étape 90 du procédé décrit en référence à la figure 2 et délivre un signal 120 correspondant au signal de voix 110 du locuteur source, mais dont les caractéristiques de spectre et de fréquence fondamentale ont été modifiées afin d'être similaires à celles du locuteur cible.

Le système décrit peut être mis en œuvre de diverses manières et notamment à l'aide d'un programme informatique adapté et relié à des moyens matériels d'acquisition sonore.

Bien entendu, d'autres modes de réalisation que celui décrit peuvent être envisagés.

Notamment, les modèles HNM et GMM peuvent être remplacés par d'autres techniques et modèles connus de l'homme de l'art, tels que par exemple les techniques dites LSF (Line Spectral Frequencies), LPC (Linear Predictif Coding) ou encore des paramètres relatifs aux formants.

REVENDICATIONS

1. Procédé d'analyse d'informations de fréquence fondamentale contenues dans des échantillons vocaux, caractérisé en ce qu'il comporte au moins :

- une étape (2) d'analyse des échantillons vocaux regroupés en trames pour obtenir, pour chaque trame d'échantillons, des informations relatives au spectre et des informations relatives à la fréquence fondamentale;

- une étape (20) de détermination d'un modèle représentant les caractéristiques communes de spectre et de fréquence fondamentale de tous les échantillons; et

- une étape (30) de détermination, à partir de ce modèle et des échantillons vocaux, d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre.

2. Procédé selon la revendication 1, caractérisé en ce que ladite étape (2) d'analyse est adaptée pour délivrer lesdites informations relatives au spectre sous la forme de coefficients cepstraux.

3. Procédé selon l'une quelconque des revendications 1 ou 2, caractérisé en ce que ladite étape d'analyse (2) comporte :

- une sous-étape (4) de modélisation des échantillons vocaux selon une somme d'un signal harmonique et d'un signal de bruit ;

- une sous-étape (5) d'estimation de paramètres de fréquence et au moins de la fréquence fondamentale des échantillons vocaux;

- une sous-étape (6) d'analyse synchronisée de chaque trame d'échantillons sur sa fréquence fondamentale; et

- une sous-étape (7) d'estimation des paramètres de spectre de chaque trame d'échantillons.

4. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce qu'il comporte en outre une étape (10) de normalisation de la fréquence fondamentale de chaque trame d'échantillons par rapport à la moyenne des fréquences fondamentales des échantillons analysés.

5. Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce que ladite étape (20) de détermination d'un modèle correspond à la détermination d'un modèle par mélange de densités gaussiennes.

6. Procédé selon la revendication 5, caractérisé en ce que ladite étape de détermination (20) d'un modèle comprend :

- une sous-étape (22) de détermination d'un modèle correspondant à un mélange de densités gaussiennes; et

- une sous-étape (24) d'estimation des paramètres du mélange de densités gaussiennes à partir de l'estimation du maximum de vraisemblance entre les informations de spectre et de fréquence fondamentale des échantillons et du modèle.

7. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que ladite étape (30) de détermination d'une fonction de prédiction est réalisée à partir d'un estimateur de la réalisation de la fréquence fondamentale sachant les informations de spectre des échantillons.

8. Procédé selon la revendication 7, caractérisé en ce que ladite étape (30) de détermination de la fonction de prédiction de la fréquence fondamentale comprend une sous-étape (32) de détermination de l'espérance conditionnelle de la réalisation de la fréquence fondamentale sachant les informations de spectre à partir de la probabilité a posteriori que les informations de spectre soient obtenues à partir du modèle, l'espérance conditionnelle formant ledit estimateur.

9. Procédé de conversion d'un signal vocal prononcé par un locuteur source en un signal vocal converti dont les caractéristiques ressemblent à celles d'un locuteur cible, comportant au moins :

- une étape (50) de détermination d'une fonction de transformation de caractéristiques spectrales du locuteur source en caractéristiques spectrales du locuteur cible, réalisée à partir d'échantillons vocaux du locuteur source et du locuteur cible; et

- une étape (70) de transformation des informations de spectre du signal de voix du locuteur source à convertir à l'aide de ladite fonction de transformation,

caractérisé en ce qu'il comporte en outre :

- une étape (60) de détermination d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre pour le locuteur cible, ladite fonction de prédiction étant obtenue à l'aide d'un procédé d'analyse selon l'une quelconque des revendications 1 à 8; et

- une étape (80) de prédiction de la fréquence fondamentale du signal de voix à convertir par l'application de ladite fonction de prédiction de la fréquence fondamentale auxdites informations de spectres transformés du signal de voix du locuteur source.

10. Procédé selon la revendication 9, caractérisé en ce que ladite étape (50) de détermination d'une fonction de transformation est réalisée à partir d'un estimateur de la réalisation des caractéristiques spectrales cibles sachant les caractéristiques spectrales source.

11. Procédé selon la revendication 10, caractérisé en ce que ladite étape (50) de détermination d'une fonction de transformation comporte :

- une sous-étape (52) de modélisation des échantillons vocaux source et cible selon un modèle de somme d'un signal harmonique et d'un signal de bruit ;

- une sous-étape (54) d'alignement entre les échantillons source et cible; et

- une sous-étape (56) de détermination de ladite fonction de transformation à partir du calcul de l'espérance conditionnelle de la réalisation des caractéristiques spectrales cibles sachant la réalisation des caractérisations spectrales sources, l'espérance conditionnelle formant ledit estimateur.

12. Procédé selon l'une quelconque des revendications 9 à 11, caractérisé en ce que ladite fonction de transformation est une fonction de transformation de l'enveloppe spectrale.

13. Procédé selon l'une quelconque des revendications 9 à 12, caractérisé en ce qu'il comporte en outre une étape (65) d'analyse du signal de voix à convertir adaptée pour délivrer lesdites informations relatives au spectre et à la fréquence fondamentale.

14. Procédé selon l'une quelconque des revendications 9 à 13, caractérisé en ce qu'il comporte en outre une étape (90) de synthèse permettant de former un signal de voix converti au moins à partir des informations de spectre transformées et des informations de fréquence fondamentale prédites.

15. Système de conversion d'un signal vocal (110) prononcé par un locuteur source en un signal vocal (120) converti dont les caractéristiques ressemblent à celles d'un locuteur cible, système comportant au moins :

- des moyens (104) de détermination d'une fonction de transformation de caractéristiques spectrales du locuteur source en caractéristiques spectrales du locuteur cible, recevant en entrée des échantillons vocaux du locuteur source (100) et du locuteur cible (102) ; et

- des moyens (114) de transformation des informations de spectre du signal de voix (110) du locuteur source à convertir par l'application de ladite fonction de transformation délivrée par les moyens (104),

caractérisé en ce qu'il comporte en outre :

- des moyens (106) de détermination d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre pour le locuteur cible, adaptés pour la mise en œuvre d'un procédé d'analyse selon l'une quelconque des revendications 1 à 8, à partir d'échantillons vocaux (102) du locuteur cible ; et

- des moyens (116) de prédiction de la fréquence fondamentale dudit signal de voix à convertir (110), par l'application de ladite fonction de prédiction déterminée par lesdits moyens (106) de détermination d'une fonction de prédiction auxdites informations de spectre transformé délivrées par lesdits moyens de transformation (114).

16. Système selon la revendication 15, caractérisé en ce qu'il comporte en outre :

- des moyens (112) d'analyse du signal de voix à convertir (110), adaptés pour délivrer en sortie des informations relatives au spectre et à la fréquence fondamentale du signal de voix à convertir ; et

- des moyens (118) de synthèse permettant de former un signal de voix converti à partir au moins des informations de spectre transformé délivrées par les moyens (114) et des informations de fréquence fondamentale prédites délivrées par les moyens (116).

17. Système selon l'une quelconque des revendications 15 et 16, caractérisé en ce que lesdits moyens (104) de détermination d'une fonction de transformation sont adaptés pour délivrer une fonction de transformation de l'enveloppe spectrale.

18. Système selon l'une quelconque des revendications 15 à 17, caractérisé en ce qu'il est adapté pour la mise en œuvre d'un procédé de conversion de voix selon l'une quelconque des revendications 9 à 12.

- des moyens (104) de détermination d'une fonction de transformation de caractéristiques spectrales du locuteur source en caractéristiques spectrales du locuteur cible, recevant en entrée des échantillons vocaux du locuteur source (100) et du locuteur cible (102) ; et

- des moyens (114) de transformation des informations de spectre du signal de voix (110) du locuteur source à convertir par l'application de ladite fonction de transformation délivrée par les moyens (104),

caractérisé en ce qu'il comporte en outre :

- des moyens (106) de détermination d'une fonction de prédiction de la fréquence fondamentale en fonction uniquement d'informations relatives au spectre pour le locuteur cible, adaptés pour la mise en œuvre d'un procédé d'analyse selon l'une quelconque des revendications 1 à 8, à partir d'échantillons vocaux (102) du locuteur cible ; et

- des moyens (116) de prédiction de la fréquence fondamentale dudit signal de voix à convertir (110), par l'application de ladite fonction de prédiction déterminée par lesdits moyens (106) de détermination d'une fonction de prédiction auxdites informations de spectre transformé délivrées par lesdits moyens de transformation (114).

16. Système selon la revendication 15, caractérisé en ce qu'il comporte en outre :

- des moyens (112) d'analyse du signal de voix à convertir (110), adaptés pour délivrer en sortie des informations relatives au spectre et à la fréquence fondamentale du signal de voix à convertir ; et

- des moyens (118) de synthèse permettant de former un signal de voix converti à partir au moins des informations de spectre transformé délivrées par les moyens (114) et des informations de fréquence fondamentale prédites délivrées par les moyens (116).

17. Système selon l'une quelconque des revendications 15 et 16, caractérisé en ce que lesdits moyens (104) de détermination d'une fonction de transformation sont adaptés pour délivrer une fonction de transformation de l'enveloppe spectrale.

18. Système selon l'une quelconque des revendications 15 à 17, caractérisé en ce qu'il est adapté pour la mise en œuvre d'un procédé de conversion de voix selon l'une quelconque des revendications 9 à 12.

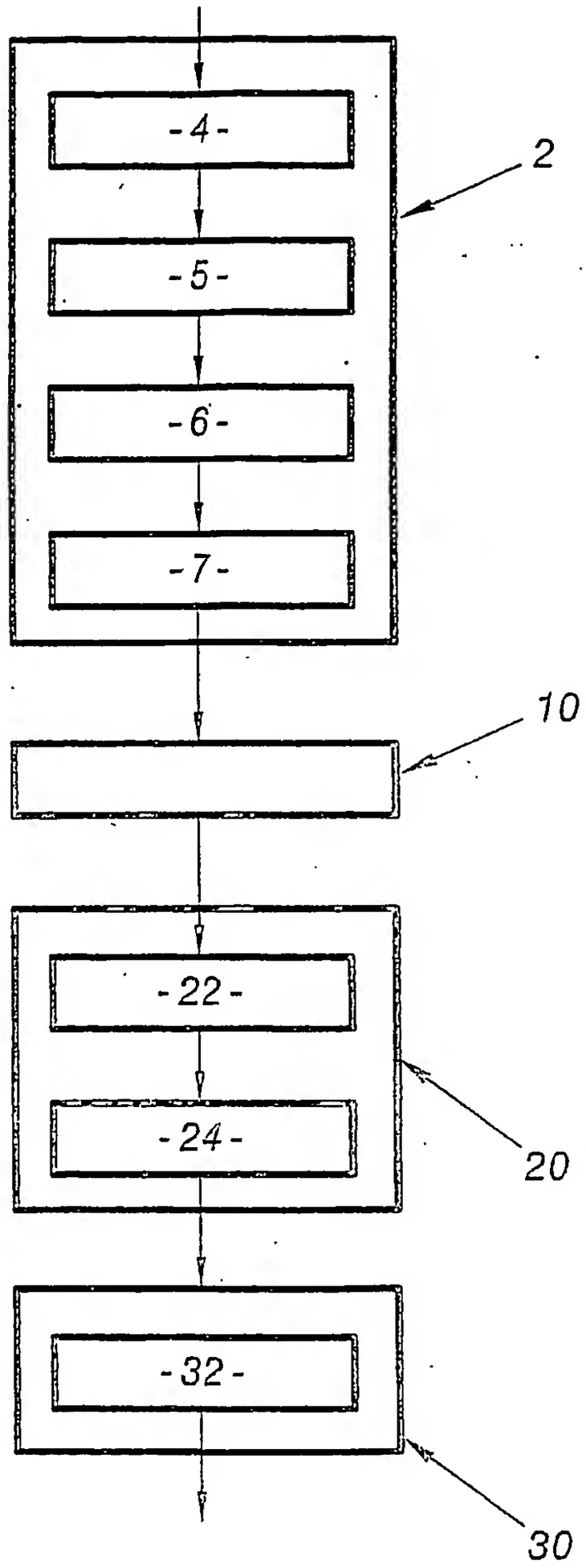

FIG. 1

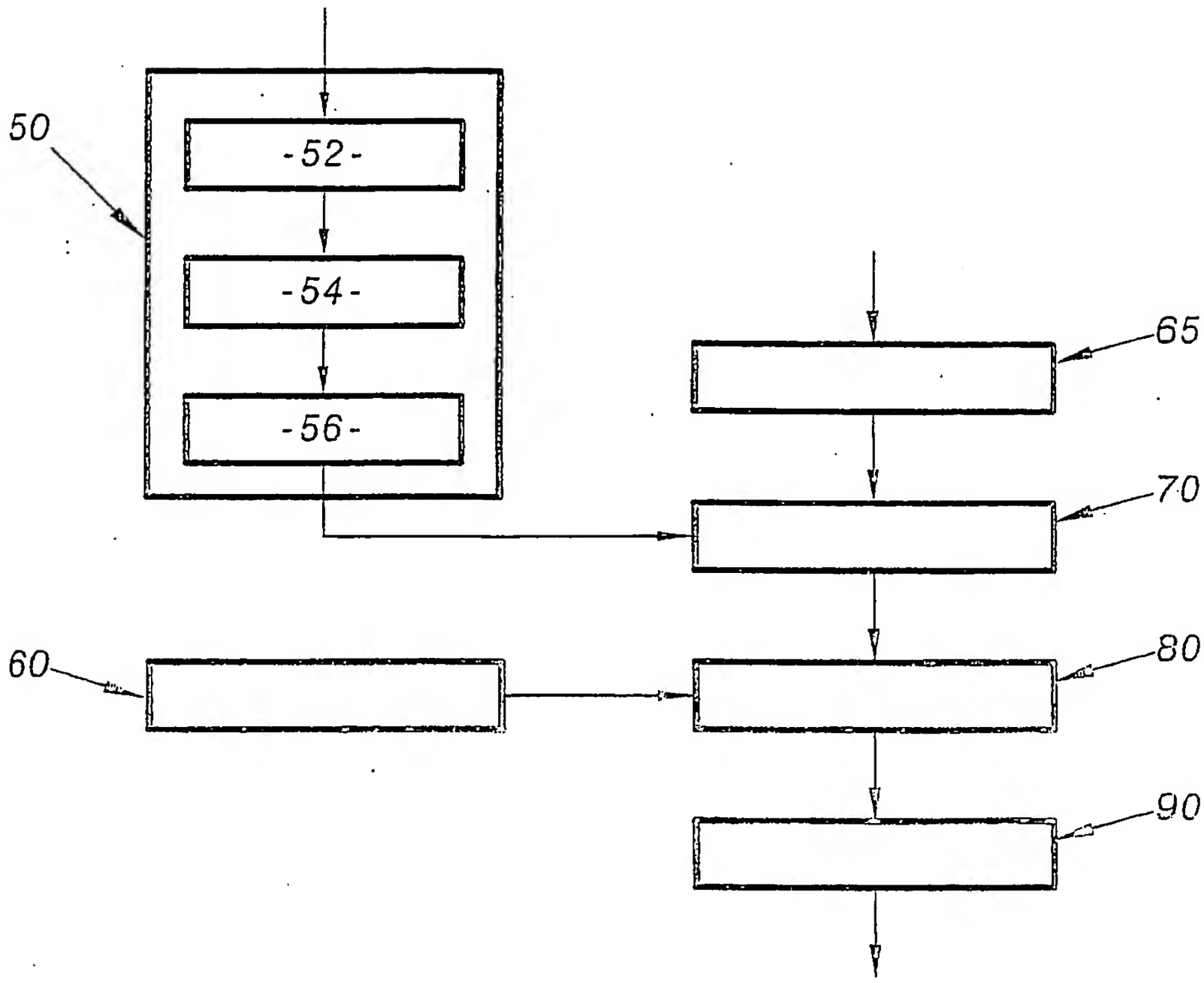

FIG.2

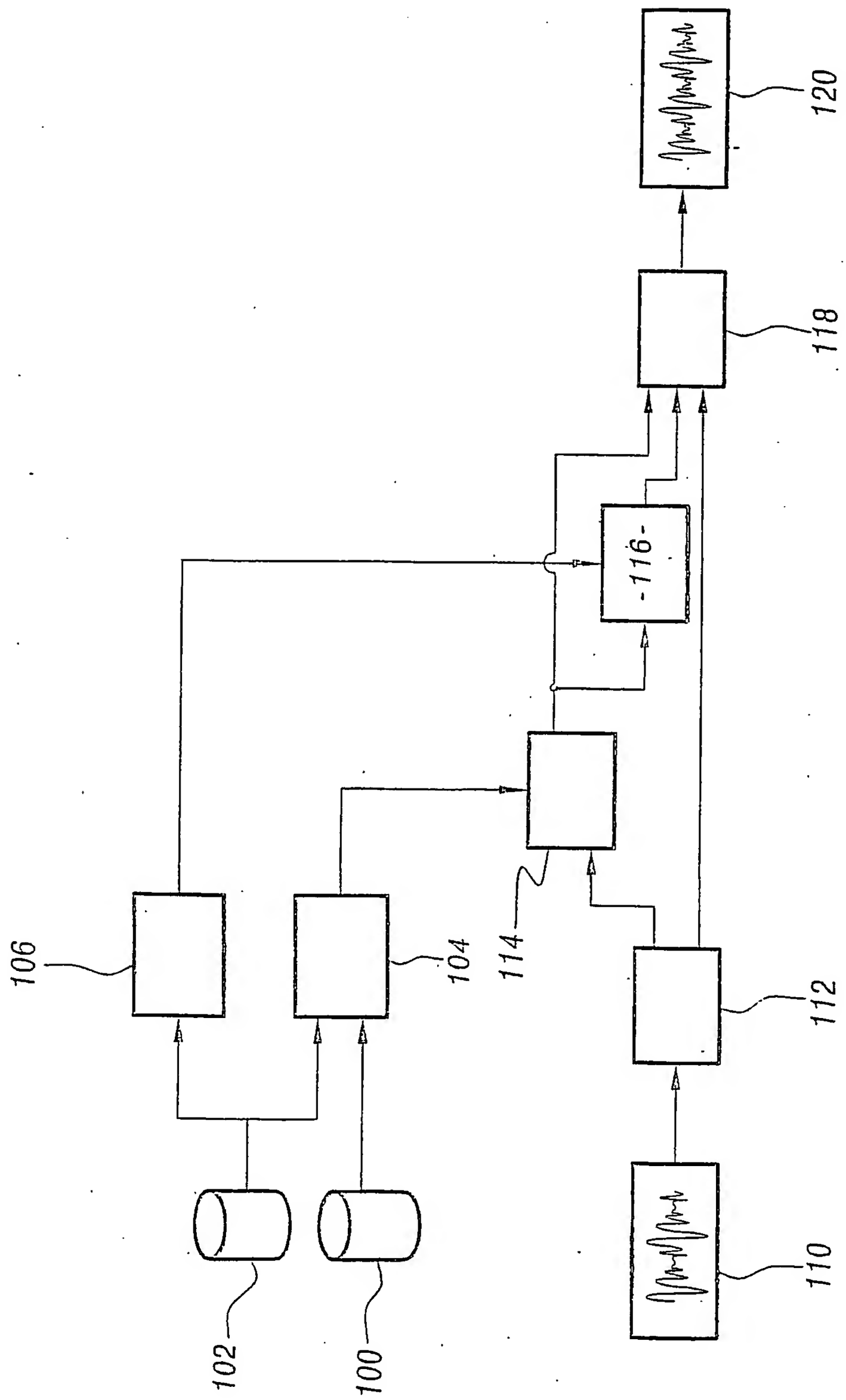

FIG.3